GERMAN AND ENGLISH WORD-ORDER REVISITED

Mark Honegger

Western Carolina University

1 Introduction

In generative grammar, word-order has typically been derived from a hierarchical syntax, and this hierarchical structure has a correspondence to the argument structure of the verb and other heads that may be altered by language-specific constraints that generate word-order. For a language like English, there has been a fairly direct relationship between word-order and the categorial distinctions represented in the syntax. Other languages, such as German, have needed more elaborate mechanisms to tweak the trees of traditional syntax to get them to correspond to the word-order patterns of the language. Andreas Kathol, in his 1995 dissertation, *Linearization-Based German Syntax*, surveys the 30 years of research on German syntax in the Generative Paradigm and finds a bewildering lack of consensus regarding the basic explanation for German word-order (Bayer and Kornfilt 1994 and Lee and Santorini 1994, for example, discuss the mixture of A and A' properties scrambled items can have in German). He then develops an alternative account which eschews a tight connection between hierarchical syntax and word-order in favor of accounting for German word directly in terms of a linear structure. If Kathol is correct in arguing that German word-order must be severed from hierarchical syntax, the theoretical consequences are far-ranging because word-order has long been one of the primary empirical phenomena to justify a hierarchical syntax in the first place. If we can't use syntax to get word-order, it questions the justification for including an autonomous syntax in generative grammar.

Kathol 1995 carries out his task of writing a linear account of German word-order by incorporating the notion from traditional German grammars of topological fields into a Head-driven Phrase Structure Grammar. The basic tenet is that German sentence patterns can be described in terms of classes of words which co-occur across different sentence types. There is no assumption that these classes share any particular syntactic value or feature. Kathol is simply trying to isolate regularities in linear placement and encode those linear positions into the lexicon and combinatorial rules of the language.

The three major sentential patterns of German are verb-final, as seen in the embedded clause in (1), verb-initial as in (2), and verb-second (3).

1) Er fragt [ob jemand den Hund füttern würde morgen abend].
he asks whether someone the dog feed would tomorrow night
'He asked if someone is going to feed the dog tomorrow night.'

2) Würde jemand den Hund füttern morgen abend?
would someone the dog feed tomorrow night
'Would someone feed the dog tomorrow night?'

3) Es kamen drei Kinder.
EXPL came three children.
'Three children came.'

Signs, that is words and phrases, in Kathol's system will come with a value called TOPO for topological field that will identify what position in the string the sign will occupy. These topological fields will be assigned to a sign either from the lexicon, via combinatorial rule or via default.

Kathol identifies three classes or fields—the *Vorfeld* ('initial field'), *Mittelfeld* ('middle field'), and *Nachfeld* ('final field'). The *Vorfeld* precedes the verb in initial or second position. The *Mittelfeld* follows the verb in initial or second position, and the *Nachfeld* follows the verb in final position. Thus, considering the embedded sentence in (1), the string *jemand den Hund* is in the *Mittelfeld* and *morgen abend* is in the *Nachfeld*. In (2), the *Mittelfeld* and *Nachfeld* are exactly the same as in (1). In (3), *es* is in the *Vorfeld* and *drei Kinder* is in the *Mittelfeld*. In addition, Kathol posits two more linear positions—a complementizer field which will be occupied either by a verb in first or second position or by a complementizer, and a verb cluster which will be occupied by one or more verbs or auxiliary elements at the end of embedded clauses.

Kathol gives the following linear precedence rule to account for the order of constituents based on their topological designation.

4) Topological LP Statement
[TOPO *vf*] < [TOPO *cf*] < [TOPO *mf*] < [TOPO *vc*] < [TOPO *nf*]
(*vf* = Vorfeld, *cf* = complementizer field, *mf* = mittelfeld, *vc* = verbal cluster, *nf* = nachfeld)

This LP rule constrains the phonological values of the signs to stand in a particular precedence relation to each other.

Let me mention a few of the mechanisms that Kathol uses in his account. First, the mittelfeld is the default location for many signs. For example, NPs are not directly assigned a topological field by the lexicon. Rather, they will by default go into the Mittelfeld unless a combinatorial rule stipulates their placement in another filed. When the last NP combines with a VP missing one of its NP arguments, the combinatory rule will stipulate that the missing argument will go into the vorfeld, or before the verb in a root indicative clause. There is no grammatical limit to the number of constituents that can appear in the mittelfeld. The only requirement in Kathol's grammar is that all the constituents in the mittelfeld be contiguous to one another. Thus, the generally free word-order of the mittelfeld is arrived at. There are some restrictions on word-order in the German mittelfeld, for example, the placement of pronouns. These can be stipulated by a fairly low level rule that is language-specific.

Unlike the mittelfeld, the vorfeld and the complementizer field allow only one constituent in their positions. Thus, competition for these positions will account for some of the regularities of German word-order. Verbs in the lexicon are stipulated to appear in either the complementizer field or the verb cluster. If a complementizer appears in the complementizer field, the verb can only appear in the verb cluster, thus arriving at the verb-final order of German embedded clauses.

Another stipulation Kathol makes is what he calls the Clause Constraint.

5) Clause Constraint
In every finite clause in German, the *cf* topological field is instantiated.

Among other things, this will insure that a verb appear in the complementizer field rather than the verb cluster if the cf is not instantiated by a complementizer or wh-word.

I applaud Kathol's basic orientation of not deriving linear order from hierarchical syntax. However, Kathol's analysis is heavily stipulatory in such a language-parochial way that it appears to give up hope of capturing any significant generalizations crosslinguistically. It depends on positing these five linear positions, which do not appear to be universally instantiated across languages. Individual fields moreover are not uniform at all but have their own idiosyncrasies. Some have a single position. Some have unlimited positions. Some are always instantiated. Some may be left empty. In particular, it is not at all clear what value an analysis based on topological fields might have for German's close cousin, English.

My proposal agrees with the spirit of Kathol's analysis in accounting for word-order on the basis of a system that has linear rather hierarchical principles. But my proposal differs from Kathol in that my system is based on principles that can be shown to be universal to all languages and that can be shown to hold for a number of linguistic phenomena besides word-order. Thus, I am trying to avoid the parochial flavor of Kathol's analysis.

2 Phonological word-order constraints

My account of English and German word order follows Anderson 1996 who proposed a set of three constraints to account for clitic placement in Serbocroatian—**EdgeMost**, **NonEdge**, and **Integrity** as seen in (6).

6) **EdgeMost(e, dir) NonEdge(e) Integrity(e)**

These constraints are ordered in an optimality framework, which relies in particular on the notion that a number of constraints hold concurrently but that these constraints are ranked so that some will win out over the others. For example, there may be a number of affixes in a language that are all realized as suffixes. Then there would be a constraint that requires them to appear on the right periphery of a word. Ranking these constraints will then determine the order of suffixes. For affixes that do not appear on the periphery of their domain but internally as an infix of some

sort, they will require two constraints, one which requires it on the periphery but also one which prohibits it from appearing first or last in its domain. Anderson suggests a cosntraint type called **Non-Edge(e)** which keeps an element e from appearing on the edge of its domain. This constraint comes in two varieties, either **Non-Initial(e)** or **Non-Final(e)**, depending on whether the element is prohibited from the initial or final edge of its domain.

A second constraint type is **EdgeMost(e, dir)** which requires an element to appear on the left or right edge of its domain. This, too, comes in two varieties, **EdgeMost(e,L)** or **EdgeMost(e,R)**. An infix will be subject to a **Non-Edge(e)** constraint and one of the **EdgeMost(e, dir)** constraints, insuring that it will appear in second or second-to-last position in its domain.

Integrity(e) disallows the placement of outside material within e. Domains are determined by the **Integrity** constraints of a language. The most typical example of this would be **Integrity(Word)** which does not allow other phonological units to occur within a word.

Apart from these constraints, lexical material can occur anywhere. That is, this theory of word-order starts out exactly opposite from the traditional theories of word-order that generative grammar has produced. Generative grammar has started from the assumption that a rigid hierarchical structure underlies word-order. Freedom in word order has been generated by mechanisms like scrambling and other types of movement that allow for more linear orders than that produced by the canonical structure of the syntax. This comes with a price—a larger theoretical apparatus. My theory starts out with the assumption that word-order is totally free. Constituents can appear anywhere in a string, even in configurational languages like English. Every language though has some constraints that do restrict the linear position of certain elements. The advantage, I propose, of looking at things this way is that there is no special theoretical mechanisms to account for free word-order languages like Warlpiri, and it turns out that it takes little theoretical apparatus to generate the more rigid word-order of languages like English. Furthermore, the constraints for both configurational and non-configurational languages turn out to be the same, i.e. both kinds of languages have the same word-order principles underlying them, and this is better than positing two different kinds of syntax for configurational and nonconfigurational languages, as many syntacticians have been forced to do (Simpson 1991).

Additionally, these constraints were first used for morphological phenomena. I argue that the same mechanisms can be used for both the placement of bound morphemes as well as the placement of free morphemes. That is, I am using the same conceptual categories as must be used for other linguistic phenomena; again that means a smaller grammar and an easier grammar to learn.

Finally, these constraints are part of the phonology of the language. They are constraints on where things should be pronounced, and this system uses the primitives that are used for other phonological processes. Phonology is sensitive to edges, to second and penultimate positions, and to the integrity of phonological material. That means, that these constraints will interact with phonological processes in the language, particularly with prosodic properties of the language. An example of this would be Heavy-NP shift in English as shown in (7).

7) Jane gave to the office a tremendously large donation that will look good on her tax return.
8) ?Jane gave to the office money.
9) Jane gave to the office MONEY (not time).

The properties that allow the direct object to appear out of its canonical position immediately after the verb are prosodic in nature. When the direct object is long enough or receives enough stress (prosodic characteristics), it appears clause-final, perhaps due to the articulatory effort required on the speaker's part.

3 German word-order

German will have a number of **Integrity** constraints that are common to all languages. Every major category in the language will have an **Integrity** constraint which will govern the close connection between the head category and its arguments, modifiers etc.

10) **Integrity(N), Integrity(P), Integrity(V),** etc.

Normally in the constraint rankings, these **Integrity** constraints will be highly ranked, but if outside material can intervene inside a phrase, there will be some other constraint(s) that are more highly ranked.

The freeness of German word-order compared to English word-order will be reflected in the lack of constraints in German which would restrict the position of an argument with respect to the verb. The relationship of all NP arguments to the verb is recovered via morphological case marking on the article and noun rather than from position in the linear string as in English. What that means is that the constituents that must be especially restricted as to location in German will be the verb, or more precisely, the verb and also the lexical item that has temporal-modal information in it, when they differ.

This latter distinction is crucial because as can be seen from root indicative clauses, the verb appears clause-final and a modal or auxiliary verb appears in second position.

11) Er ist nach Hause gegangen.
He has gone home.

The same kind of ordering considerations will be seen to apply to English in a moment as well. As I have proposed for English, there is in German a category which I will label TM for time-modality element. This is akin to the semantic counterpart of the syntactic category Infl. It carries information pertaining to tense and modality.

This also means that some words in the lexicon will be encoded in terms of this TM element. For German and English, both modals and finite verbs will have a time-modality element. Finite verbs, in particular, are portmanteaus, a combination of a verb and a time-

modality element at the same time. When they are subject to two or more constraints, the highest one takes precedence.

German has these constraints for the time-modality element.

12) a.	**EdgeMost($TM_{Interrogative}$,L)**	**[TM1]**
b.	**EdgeMost($TM_{Imperative}$,L)**	**[TM2]**
c.	**EdgeMost($TM_{non\text{-}illocutionary}$,R)**	**[TM3]**
d.	**Non-Initial($TM_{indicative}$) >> EdgeMost($TM_{Indicatives}$,L)**	**[TM4] & [TM5]**

Together, all of these constraints capture the phenomenon that in German, the location of the word bearing the time-modality element determines the illocutionary force of the clause. The constraints in (12a) tell us that when the lexical item bearing the time-modality element occurs first in the sentence, the sentence is encoded with interrogative or imperative force. The constraints in (12d) identify sentences with a lexical item bearing a time-modality element in second position. Those sentences will be encoded with indicative force.

The constraint in (12c) tells us that a clause with a lexical item bearing a time-modality element in final position will be encoded as having no illocutionary force. This accounts for the verb-final order of subordinate clauses in German. Subordinate clauses have no illocutionary force. By themselves, they do not assert, question or command. So the problem with (13) is a semantic problem.

13) *Ich glaube, daß Barbara ist heute zu Hause.

The subordinate conjunction *daß* requires a clause without illocutionary force, yet the time-modality element in (13) signals it as an indicative clause. Therefore, the sentence is ungrammatical for semantic reasons, not any violation of syntax and not any competition for a single position, whether that position occurs in a hierarchical syntax a la minimalism/Principles & Parameters or a linear system a la Kathol's topological fields.

Likewise, this constraint explains why root clauses cannot have final lexical items with time-modality element.

14) *Er nach Hause geht.

In (14), the constraints listed in (12) identify the sentence as having no illocutionary force. It makes no assertion or question or command and thus is semantically anomalous.

There is one constraint that must be added to my description of German to account for the behavior of separable prefixes. As is well-known, German has a set of separable prefixes that occur as prefixes on verbs yet can never appear in second position, as the following sentences illustrate.

15) *Peter zurückkommt morgen aus dem Urlaub.
16) Peter kommt morgen aus dem Urlaub zurück.

17) Peter wird morgen aus dem Urlaub zurückkommen.

These sentences require the following constraints and ordering.

18)
TM1 >>
TM2 >> EdgeMost(V,R) < EdgeMost(SepPrefix,R) >> Integrity(SepPrefix/Verb)
TM3 >>
TM4 >> TM5 >>

The ordering in (18) has the **EdgeMost** constraint outranking the **Integrity** constraint for the verb and its separable prefix. This means that no matter where the verb might appear in the sentence, the separable prefix must be on the right periphery of its clause. All the time-modality element constraints are ranked higher than the constraints related to the verb and its prefix. That means that the TM constraint may require the lexical item carrying the TM, which is sometimes a verb, to appear elsewhere in the sentence. Notice what happens, then, if we have a subordinate clause or a root clause with a modal. In a subordinate clause, the **EdgeMost** constraint governing the verb or element with the TM outranks the **EdgeMost** constraint governing the separable prefix. This ensures that even though the separable prefix must appear on the right edge of its domain, it will still appear as a prefix on the verb when the verb is clause-final and not as a suffix. In an embedded clause with a modal, **TM3** which requires the time-modality element to appear on the right edge of the clause outranks **EdgeMost(V,R)** which outranks **EdgeMost(SepPrefix,R)** , thus giving us the correct order for a sentence like (19).

19) Ich glaube daß Barbara heute zu Hause muß.
I believe that Barbara must be at home today.

Likewise, in a root clause with a modal, the TM constraints will require the modal to appear in either first or second position in the clause because the modal carries the time-modality element. This leaves the verb to occur last in its clause, after the separable prefix, because **EdgeMost(V,R)** outranks **EdgeMost(SepPrefix,R)**.

4 English word-order

Let's compare this analysis of German word-order to what is needed to account for basic English word-order. The constraints that are used in this paper for word-order are also those that govern morphological phenomena as well, but I will leave out those that are morphological for the sake of simplicity in this paper. The basic ordering is in (20).

20) **Non-Initial($TM_{indicative}$) >> EdgeMost($TM_{indicative}$,L) >> EdgeMost(Verb,L)**
EdgeMost($TM_{interrogative}$,L) >>
EdgeMost($TM_{imperative}$,L) >>

To these must be added principles that identify the arguments of the verb.

First, I must define an immediate domain. The immediate domain is the set of arguments and modifiers that belong to any category that takes arguments and modifiers. Nouns in English, for example, will be a part of the immediate domain of verbs and prepositions. The immediate domain which a noun appears in is determined by the major lexical categories the noun is adjacent to as they are encountered in linear order. For example, in (21)

21) The boy who was on top of the log ate the hot dog.

the constituent *the log* would not be identified as the subject of *ate* rather than the object of the preposition *of*, because *of* appears first in the string. Therefore, *of* must identify its domain first, and any noun following the preposition will be identified as its object.

Arguments, then are identified as per (22).

22) For all Ns in the immediate domain of V:
a. The subject is to the left of V.
b. The direct object is the rightmost N in the domain.
c. The indirect object is between V and the direct object.

Let me give a few sentence to illustrate how these constraints interact in English. In (23),

23) The child kicked the ball.

the word *kicked* is entered in the lexicon as a portmanteau consisting of a time-modality element and a verb. The $TM_{indicative}$ constraints in (20) insure that *kicked* will appear second in its clause. The noun that appears before *kicked* will be identified as the subject and the one following *kicked* will be identified as the object. If the verb appeared first as in (24),

24) *Kicked the child the ball.

the sentence would violate both **Non-Initial($TM_{indicative}$)** and the argument identification principle in (22a).

For a question like (25),

25) Did the child kick the ball?

EdgeMost($TM_{interrogative}$,L) constrains the word *did*, which has the time-modality element in this sentence, to appear on the left edge of the clause. The subject *the child* and the direct object *the ball*, then appear on the correct edges of the verb to have their argument status identified correctly. This analysis shows *do*-support to be a means of allowing English to satisfy argument identification principles (subject must be to the left of the verb) and constraints on time-modality elements (TM must be on the left of its domain) simultaneously (Honegger 1997a,b). The principles of English word-order cannot be satisfied in interrogatives unless the verb is separated from the time-modality element.

The argument identification principles in (22) are one of the major differences between German and English word-order. Note though where the constraints for German and English are similar or different. The constraints for the time-modality elements are similar except the clause-final constraint for TM in German. The other singular difference is the **EdgeMost** constraint for verbs. English has an **EdgeMost(Verb,L)** and German an **EdgeMost(Verb,R).**

5 Conclusion

In conclusion, there are two things that need to be drawn to the reader's attention at this point. One is that the theory I am proposing will not account for all word-order facts. For example, the kind of ungrammaticality resulting from superiority effects as in (26) are not explained by my constraint-based system.

26) *What did John say who ate?

The kind of system I have presented today will allow such strings. These kinds of ungrammaticality will need a different kind of explanation, one I would give on the basis of semantics, not phonology.

Second, there are certain kinds of explanations that have been traditionally used to explain some phenomena that are not available to my theory, in particular the idea of certain constituents competing for the same position in the syntax.

Both Kathol and GB/Minimalist theories rely on this explanation for German. In Kathol's analysis, the Vorfeld and the Complementizer Field both contain only one position—one constituent at most cann appear in either place.

In GB/Minimalism, there is a single Spec node for each CP. The upshot for both theories is that if one consituent appears in the CF for Kathol or Spec of CP for the Minimalist syntactician, it precludes another constituent from appearing there. Thus, there is the assertion that there is a complementarity between subordinate clauses and root clauses in German. In the former, the complementizer occurs in Spec of CP. In the latter, the verb does. Kathol uses this kind of reasoning to explain why a clause cannot have both an initial wh-phrase and complementizer as in (27) for Standard German.

27) *Ich frage mich [wen daß Adam gesehen hat].

Again, this kind of explanation is not available under my system where there are no positions based on either a linear or hierarchical structure. If there is a competition between *wen* and *daß* in (27), it would have to be a semantic competition.

REFERENCES

Bayer, Josef, and Jaklin Kornfilt. 1994. Against scrambling as an instance of Move-alpha. Studies on Scrambling: Movement and non-movement approaches to free word-order phenomena, ed. by Norbert Corver and Henk van Riemsdijk, 17-60. New York: Mouton de Gruyter.

Honegger, Mark. 1997. A phonological account of the "Adverb Effect" and *THAT*-t violations. 1996 Mid-America Linguistics Conference Papers, ed. by Clifton Pye, 168-80. Lawrence: University of Kansas.

———. 1997. A unified account of configurational and non-configurational languages. 33rd Regional Meeting of the Chicago Linguistics Society.

Kathol, Andreas. 1995. Linearization-based German syntax. Ohio State Ph.D. dissertation in linguistics.

Lee, Young-Suk, and Beatrice Santorini. 1994. Towards resolving Webelhuth's paradox: Evidence from German and Korean. Studies on Scrambling: Movement and non-movement approaches to free word-order phenomena, ed. by Norbert Corver and Henk van Riemsdijk, 257-300. New York: Mouton de Gruyter.

Simpson, Jane. 1991. Warlpiri morpho-syntax: A lexicalist approach. Kluwer: Dordrecht.